

Zeitaggregation und Ereignisdaten: eine Simulation zu den Auswirkungen der Prozeßzeitskalierung

Steinhage, Nikolei

Veröffentlichungsversion / Published Version

Arbeitspapier / working paper

Empfohlene Zitierung / Suggested Citation:

Steinhage, N. (2000). *Zeitaggregation und Ereignisdaten: eine Simulation zu den Auswirkungen der Prozeßzeitskalierung*. (Arbeitspapier / Sfb 186, 66). Bremen: Universität Bremen, SFB 186 Statuspassagen und Risikolagen im Lebensverlauf. <https://nbn-resolving.org/urn:nbn:de:0168-ssoar-57610>

Nutzungsbedingungen:

Dieser Text wird unter einer CC BY-NC-ND Lizenz (Namensnennung-Nicht-kommerziell-Keine Bearbeitung) zur Verfügung gestellt. Nähere Auskünfte zu den CC-Lizenzen finden Sie hier:

<https://creativecommons.org/licenses/by-nc-nd/4.0/deed.de>

Terms of use:

This document is made available under a CC BY-NC-ND Licence (Attribution-Non Commercial-NoDerivatives). For more information see:

<https://creativecommons.org/licenses/by-nc-nd/4.0>

**Sonderforschungsbereich 186
der Universität Bremen**

**Statuspassagen und Risikolagen
im Lebensverlauf**

Zeitaggregation und Ereignisdaten

**Eine Simulation zu den Auswirkungen der
Prozeßzeitskalierung**

von

Nikolei Steinhage

Arbeitspapier Nr. 66
Februar 2000



Zeitaggregation und Ereignisdaten

Eine Simulation zu den Auswirkungen der
Prozeßzeitskalierung

von

Nikolei Steinhage

Bremen, den 11. Februar 2000

Vorwort

Der Schwerpunkt des Teilprojekts Haushaltsdynamik und soziale Ungleichheit im internationalen Vergleich B6 liegt nicht nur auf der Untersuchung von Haushaltsdynamik und sozialer Ungleichheit im Lebenslauf, sondern auch auf der Weiterentwicklung von quantitativen Methoden zur Analyse von Längsschnittdaten. Zu solch einem Vorhaben bedarf es der Ausleuchtung der genutzten Methoden, einer Herausarbeitung von Schwächen und Problemen, die mit den genutzten Methoden und den zu analysierenden Daten verbunden sind, um daran ansetzend Lösungen zu erarbeiten.

Der vorliegende Beitrag soll das in der empirischen Forschung relativ unbeachtet gebliebene Problem der Zeitaggregation und des damit oftmals einhergehenden Short-Spell-Problems bei der Analyse von Ereignisdaten verdeutlichen. In Bezug auf die Zeitaggregation scheint die Konzeption diskreter Ereignisanalyse als ein gleichsam natürliches Verfahren des Umgangs mit diesem Problem zu sein. Es wird gezeigt, dass die zu behandelnde Frage weniger ein methodisches, als vielmehr ein Datenproblem ist, welches auch nicht mit den Methoden diskreter Ereignisanalyse gelöst werden kann. Die hier behandelten Probleme beschränken sich nur – wie oftmals angenommen – auf den abhängigen Prozeß, sondern auch auf parallele und interdependente Prozesse.

Bremen, den 11. Februar 2000

Prof. Dr. Walter R. Heinz

Sprecher des Sonderforschungsbereichs 186

Inhaltsverzeichnis

1	Einleitung	1
2	Selektion und Zeitaggregation	1
3	Diskrete Modelle und Zeitaggregation	4
4	Beispiele	7
4.1	Daten	8
4.2	Simulation von Beobachtungsintervallen	9
4.3	Exponentialmodelle	11
4.4	Periodenspezifische Exponentialmodelle	16
4.5	Competing-Risk Exponentialmodelle	20
4.6	Diskussion	25
5	Zusammenfassung	26
6	Literaturverzeichnis	26

Abbildungsverzeichnis

1	Simulation von Beobachtungsintervallen	10
2	Survivor-Funktionen für verschiedene Zeitaggregationen . . .	15

Tabellenverzeichnis

1	Stetiges Exponentialmodell und diskrete Exponentialmodelle mit Simulationsintervallen von einem, 12, 24, 48 und 96 Monaten	13
2	Periodenspezifisches Exponentialmodell für ein stetiges Modell und für Simulationsintervalle von einem und 12 Monaten . . .	18
3	Periodenspezifisches Exponentialmodell für Simulationsintervalle von 24, 48 und 96 Monaten	19
4	Competing-Risk Exponentialmodell für ein stetiges Modell mit Splits auf Monatsbasis	21
5	Competing-Risk Exponentialmodell für Simulationsintervalle von einem Monat	21
6	Competing-Risk Exponentialmodell für Simulationsintervalle von 12 Monaten	22
7	Competing-Risk Exponentialmodell für Simulationsintervalle von 24 Monaten	22

8	Competing-Risk Exponentialmodell für Simulationsintervalle von 48 Monaten	24
9	Competing-Risk Exponentialmodell für Simulationsintervalle von 96 Monaten	24

1 Einleitung

Bei der Schätzung von Parametern in ereignisanalytischen Modellen wird normalerweise angenommen, daß die Zeit kontinuierlich gemessen worden ist oder daß Ereignisse nur zu diskreten Zeitpunkten auftreten können beziehungsweise die Rate zwischen zwei Zeitpunkten konstant ist. Das Eintreten von Ereignissen in der Zeit kann nicht exakt oder kontinuierlich gemessen werden. Auch treten Ereignisse in den meisten Fällen nicht zu den diskreten Zeitpunkten auf, zu denen sie gemessen worden sind. Es wird daher angenommen, die Ereignisse seien zeitlich genau genug gemessen, um als Approximation an einen kontinuierlichen Prozeß fungieren zu können, oder daß die Ereignisse tatsächlich zu diskreten Zeitpunkten auftreten. In den Sozialwissenschaften liegen Längsschnittdaten meist nur auf Monats- oder Jahresbasis genau vor. Ob auf einer solchen zeitlichen Basis die oben erwähnten Bedingungen als (annähernd) erfüllt betrachtet werden können, hängt von den Eigenschaften (der Dynamik, der Geschwindigkeit) der zu untersuchenden Prozesse ab.

Das hier zu behandelnde Problem ist nun folgendes. Forscher wissen normalerweise, daß die Dauer, die eine Analyseeinheit in einem Zustand verbracht hat, zwischen $T - 1$ und T liegt, nicht aber die exakte Dauer. Praktisch wird daher zumeist T als die wirkliche Dauer angenommen, was zu einem Bias in den Parameterschätzungen führt, der Zeitaggregationsbias genannt werden kann.

In diesem Beitrag soll die Problematik diskutiert und anhand eigener Analysen verdeutlicht werden, die sich ergibt, wenn zu sorglos mit dem Zeitaggregationsproblem umgegangen wird. In diesem Zusammenhang ist darauf zu verweisen, daß sich aus der Konzeption diskreter ereignisanalytischer Methoden leicht ein gewisses Vertrauen ergibt, mit Ereignissen umgehen zu können, die zeitlich ungenau gemessen worden sind. Inwiefern und unter welchen Bedingungen sich dieses Vertrauen als gerechtfertigt erweist, wird in diesem Beitrag untersucht.

2 Selektion und Zeitaggregation

Obwohl das Problem des Zeitaggregationsbias bekannt ist, mögen Forscher glauben, daß dieses zu vernachlässigen sei und die inkonsistenten Schätzungen sich nicht wesentlich von den konsistenten unterscheiden. Dieses mag unter bestimmten Umständen zutreffen. Entsprechende Umstände lassen sich allerdings nicht voraussehen und aus den vorliegenden Daten auch nicht ableiten. Im Folgenden soll das Problem des Zeitaggregationsbias zusammen mit dem Short-Spell-Problem, welches mit diesem meist einhergeht, näher dargestellt werden. Bei diesem zweiten Problem handelt es sich um einen Selektionsmechanismus, der aus dem Beobachtungsdesign selbst resultiert.

Wenn wir wissen, ab wann eine Analyseeinheit im Zustand j war, und wir wissen, daß sie sich zum Zeitpunkt $t - 1$ immer noch in diesem Zustand befand, zum Zeitpunkt t aber nicht mehr, was ist dann die optimale Wahl der Dauer T , in der sich die Analyseeinheit im Zustand j befand? Das Problem ist nicht so trivial, wie es zunächst erscheinen mag. Die Bestimmung von T hat Auswirkungen auf die Parameterschätzungen, kann den Zeitverlauf der Rate verändern und macht es unter Umständen problematisch, eine zeitliche Ordnung zwischen erklärenden Variablen und der Rate herzustellen.

Wenn wir wissen, daß ein Ereignis zwischen $t - 1$ und t stattfand, also im Zeitfenster Δt , und wir t als den wahren Zeitpunkt des Eintritts des Ereignisses behandeln, so werden wir die eigentliche Dauer bis zum Eintritt des zu untersuchenden Ereignisses überschätzen. Je größer Δt , desto größer ist der Schätzfehler. Durch die Überschätzung der tatsächlichen Episodendauern scheint der Prozeß langsamer abzulaufen, als es in Wirklichkeit der Fall ist. Würden wir $t - 1$ als den wahren Zeitpunkt des Eintritts des Ereignisses wählen, so würden wir die Episodendauern unterschätzen und der Prozeß würde schneller ablaufen.

Desweiteren ist der Bias in den Schätzungen umso größer, je höher die Rate $r(t)$ ist. D.h. je kürzer die Verweildauer T , desto höher die Rate $r(t)$ und desto größer der Bias. Dies ergibt sich daraus, daß nicht nur die Größe von Δt den Bias beeinflusst, sondern auch die Verteilung der Episodendauern im Zeitfenster Δt . Je höher die Rate $r(t)$, desto kürzer werden die Episodendauern im Durchschnitt sein. Wie groß der Schätzfehler tatsächlich sein wird, hängt wiederum von der Wahl des Ereigniszeitpunktes ab. Das Problem dabei ist, daß sich anhand der zu analysierenden Daten nicht sagen läßt, ob $t - 1$, t oder ein Wert dazwischen eine Wahl darstellt, die zu einer Minimierung des Schätzfehlers führt.

In einer Arbeit hat Petersen (1991) gezeigt, daß, bei Nichtkenntnis der Verteilung der Episodendauern im Zeitfenster Δt , die optimale Wahl des Zeitpunktes des Ereignisses zwischen $t - 1$ und $t - 0.5$ liegt, da wir die Rate, also die Geschwindigkeit des Prozesses, nicht kennen und es mit zunehmender Größe von Δt wahrscheinlicher wird, daß das Ende der Episode in der ersten Fensterhälfte (für $T = t$ also zwischen $t - 1$ und $t - 0.5$) liegt.

Wird eine Verweildauerabhängigkeit der Rate mit in die Überlegungen einbezogen, so verkompliziert sich die Situation zusätzlich. Ist d der Wert zwischen $t - 1$ und t , welcher den Bias minimiert, so muß d bei einer positiven Verweildauerabhängigkeit steigen und bei einer negativen sinken, da die Verteilung der Ereignisse im Zeitfenster Δt nicht gleich ist.

Neben dem Zeitaggregationsproblem tritt häufig ein weniger beachtetes Problem auf, welches sich aus dem Beobachtungsdesign selbst ergibt und das Short-Spell-Problem genannt werden kann. Das Problem resultiert daraus, daß nur Veränderungen in Zuständen von einem Beobachtungszeitpunkt zu einem anderen beobachtet werden. Gab es mehr als einen Wechsel zwischen

zwei Beobachtungszeitpunkten, so können dies bei bestimmten Erhebungsdesigns (z.B. einer Panelerhebung ohne retrospektive Fragen) nicht in Erfahrung gebracht werden. Mit zunehmender Größe von Δt steigt also die Wahrscheinlichkeit an, Episoden, die kürzer als Δt sind, nicht mehr beobachten zu können.¹

Die Wahrscheinlichkeit, eine Episode mit einer Dauer von mindestens Δt zu beobachten, ist gleich eins. Ist die Episodendauer T jedoch kürzer als Δt (oder gleich), so beträgt die Wahrscheinlichkeit:

$$P = 1 - \frac{\Delta t - T}{\Delta t}, \quad T \leq \Delta t. \quad (1)$$

Die Zeitaggregation führt dazu, daß die Varianz von T der zeitaggregierten Daten geringer ist, als die der wirklichen T . Der Effekt des Short-Spell-Problems verstärkt die Unterschiede in den Varianzen weiter. Zusätzlich erhöht sich die durchschnittliche Verweildauer. Beide Effekte, Zeitaggregation und Short-Spells, weisen also in die gleiche Richtung. Was sich durch das Short-Spell-Problem (und auch durch die Zeitaggregation) ändert, ist die Verteilung der Episodendauern und damit auch der Zeitverlauf der Rate. So kann beispielsweise aus einer zunächst ansteigenden und dann abfallenden Rate aufgrund des Short-Spell-Problems eine abfallende Rate geschätzt werden, da die Ereignisse, die für den anfänglichen Anstieg der Rate verantwortlich sind, aufgrund der Größe von Δt weniger häufig beobachtet werden und, wenn sie beobachtet werden, ihre Dauer überschätzt wird.

Desweiteren kann aufgrund des Selektionsproblems kurzer Episoden nicht mit Sicherheit gesagt werden, wie der Zustandswechsel in einer Variablen verlaufen ist, beispielsweise eine Veränderung im Berufsprestige durch Jobwechsel. Auch ist oftmals anzunehmen, daß der Einfluß unabhängiger Variablen auf die kurzen Episoden ein anderer ist als auf die langen.

Gerade die Berücksichtigung von zeitveränderlichen Kovariablen im Hinblick auf die Herstellung einer kausalen Ordnung zwischen Prozessen wird als eine Stärke ereignisanalytischer Methoden angesehen. Das Nichtwissen um den genauen Zeitpunkt des Eintritts des Ereignisses zwischen $t - 1$ und t hat aber Folgen für die Etablierung einer kausalen Ordnung zwischen einer unabhängigen Variable X und der abhängigen Variable Y , denn als eine Voraussetzung einer kausalen Ordnung muß eine zeitliche Ordnung gegeben sein (Kelly und McGrath 1988). Bei zeitveränderlichen Variablen ist es ebenfalls unbestimmt, wann in dem Fenster Δt ein Zustandswechsel in X stattgefunden hat, und, wenn in Y in diesem Fenster ebenfalls ein Zustandswechsel stattgefunden hat, ob dieser zeitlich vor oder nach dem in X erfolgt ist. Zeitveränderliche erklärende Variablen sind auch Prozesse. Eine Zeit lang nimmt eine Analyseeinheit einen bestimmten Zustand ein und wechselt dann

¹Beispielsweise wird in manchen Panelerhebungen nur nach einem Jobwechsel im letzten Jahr gefragt, wodurch weitere Jobwechsel unberücksichtigt bleiben.

eventuell in einen anderen Zustand. Da dieser parallele Prozeß ebenfalls von einer Zeitaggregation betroffen sein kann, können besonders Variablen, die sich auf zeitliche Dimensionen beziehen (Alter, Erfahrung), nicht mehr genau angegeben werden, da sie sich während des Intervalls Δt ändern und daher ihr Einfluß nicht genau genug im Modell berücksichtigt werden kann. Problematisch wird dies besonders dann, wenn der Einfluß dieses unabhängigen Prozesses keinen linearen Einfluß auf den abhängigen Prozeß hat.

Zusammenzufassen ist folgendes: Das Ausmaß eines Bias in den Parameterschätzungen, hervorgerufen durch das Zeitaggregationsproblem und des damit oftmals einhergehenden Short-Spell-Problems, läßt sich aus den zu analysierenden Daten nicht ableiten. Beide Probleme gelten nicht nur für den abhängigen Prozeß, sondern auch für parallele Prozesse (zeitveränderliche erklärende Kovariablen). Desweiteren kann aufgrund der Zeitaggregation unter bestimmten Umständen keine zeitliche Ordnung zwischen Prozessen ausgemacht werden.

3 Diskrete Modelle und Zeitaggregation

Aus der Konzeption diskreter ereignisanalytischer Modelle, die annehmen, daß Ereignisse nur zu diskreten Zeitpunkten auftreten können oder/und gemessen werden, ergibt sich gegenüber stetigen Modellen leicht ein gewisses Vertrauen, das Problem des Aggregationsbias wenn nicht zu lösen, so doch zumindest vermindern zu können. Ob dieses Vertrauen zu rechtfertigen ist, soll, nachdem die Konzeption diskreter Ereignisanalyse dargestellt worden ist, näher erörtert werden.

Stetige Übergangsratenmodelle beschreiben diskrete abhängige Variablen, die ihren Zustand zu jedem Zeitpunkt verändern können. Der Zustandsraum ist also diskret und die Zeitachse kontinuierlich. Ein solches Modell beschreibt den Prozeß zu jedem Zeitpunkt. Das grundlegende Übergangsratenmodell lautet wie folgt (vgl. u.a. Blossfeld und Rohwer 1995: 31; Blossfeld, Hamerle und Mayer 1986: 31):

$$r(t) = \lim_{\Delta t \rightarrow 0} \frac{P(t \leq T < t + \Delta t | T \geq t)}{\Delta t}. \quad (2)$$

T ist hier eine Zufallsvariable, welche die nicht zensierte Dauer bis zum Eintreten eines Ereignisses angibt. Die Übergangsrate $r(t)$ ist der Grenzwert der bedingten Wahrscheinlichkeit, daß die Episode im Intervall $[t, t + \Delta t]$ zu Ende geht, unter der Voraussetzung, daß die Episode bis zum Beginn des Intervalls andauert. Die Werte der Rate selbst sind jedoch keine Wahrscheinlichkeiten, da sie größer als Eins sein können (Blossfeld, Hamerle und Mayer 1986: 32). Sie sollten daher als eine Neigung interpretiert werden. Für ein sehr kleines

Δt kann $r(t)\Delta t$ aber als Approximation der bedingten Wahrscheinlichkeit $P(t \leq T < t + \Delta t | T \geq t)$ aufgefaßt werden.

Obwohl stetige Zeitmodelle plausiblere Repräsentationen der Ereignisse generierenden Prozesse sind, kann Zeit immer nur diskret gemessen werden, auch wenn die Messung in noch so kleinen Einheiten erfolgt. Wenn die diskreten Zeiteinheiten relativ klein sind, dann ist es akzeptabel ihre Diskretheit zu ignorieren und ihre Messung als eine kontinuierliche anzunehmen.

Die Notation diskreter Zeitmodelle ist die der stetigen sehr ähnlich. Es wird angenommen, daß die Zeit nur positive Ganzzahlen annehmen kann ($t_i = 1, 2, 3, \dots$), wobei gilt, daß $t_1 < t_2 < t_3 \dots$. Dann wird begonnen, insgesamt n voneinander unabhängige Individuen ($1, \dots, n$) zu dem Zeitpunkt $t = 1$ zu beobachten. Die Beobachtung jedes Individuums wird weitergeführt bis zum Zeitpunkt t_i . Zu diesem Zeitpunkt fand entweder ein Ereignis statt oder die Beobachtung ist rechtszensiert. Es wird gewöhnlich angenommen, daß der Zeitpunkt der Zensierung unabhängig von der Rate ist. Die Definition einer zeitdiskreten Rate ist

$$\lambda(t_i) = P(T = t_i | T \geq t_i), \quad (3)$$

wobei T eine diskrete Zufallsvariable ist, die die nicht zensierte Dauer bis zum Auftreten eines Ereignisses angibt. Dieses ist die bedingte Wahrscheinlichkeit, daß ein Ereignis zum Zeitpunkt t_i eintritt, unter der Bedingung, daß es nicht schon eingetreten ist (vgl. Allison 1982).

Der nächste Schritt ist es, zu spezifizieren, wie die Rate von der Zeit und den erklärenden Variablen abhängt. Die häufigste Wahl der Parametrisierung der bedingten Wahrscheinlichkeit ein Ereignis zu erfahren in Abhängigkeit von Zeit und Kovariaten ist die Funktion der logistischen Regression (Agresti 1990: 90; Allison 1982; Diggle, Liang und Zeger 1994):

$$P(Y = 1 | X_i, X_k(t)) = \frac{\exp(\alpha_t + \sum_i \beta_i x_i + \sum_k \beta_k x_k(t))}{1 + \exp(\alpha_t + \sum_i \beta_i x_i + \sum_k \beta_k x_k(t))}, \quad (4)$$

wobei α_t die Konstante ist und, wenn die Rate als zeitveränderlich modelliert wird, den Zeitverlauf angibt. Ferner stellen X_i zeitkonstante Variablen und $X_k(t)$ zeitveränderliche unabhängige Variablen dar.

Gegenüber der stetigen Ereignisanalyse besitzt die diskrete Ereignisanalyse einige Vorteile, die ihre Anwendung einfacher gestalten. Zum einen ist es vollkommen unproblematisch, zeitveränderliche unabhängige Variablen in ein Modell mit einzubeziehen. Zum Zweiten können bei der Schätzung diskreter Modelle bereits bekannte und gut verstandene Methoden verwendet werden. Ein Grund, der vor wenigen Jahren, als Computer teuer und nicht so leistungsstark waren, noch zählte, war die Möglichkeit, diskrete Modelle durch

log-lineare Methoden zur Analyse von Kontingenztabellen zu berechnen, wodurch sich der für die Analysen nötige Zeitaufwand verringern ließ. Da aber mittlerweile genügend und günstige Rechenkapazitäten zur Verfügung stehen und auch die Einbeziehung von zeitveränderlichen, unabhängigen Variablen in stetige Modelle durch die Methode des Episodensplittings (vgl. Blossfeld und Rohwer 1995: 120f) kein Problem mehr darstellen sollte, fallen diese Vorteile weg. Es bleibt also nur die Anwendbarkeit bekannter und in vielen Statistikprogrammen implementierter Methoden als ein Vorteil bestehen.

Unter bestimmten Bedingungen mag es aber trotzdem von Vorteil sein, einem diskreten Modell den Vorzug gegenüber einem kontinuierlichen zu geben. Dieses ist dann der Fall, wenn Ereignisse tatsächlich nur zu diskreten Zeitpunkten auftreten oder gemessen werden können (Allison 1982: 63). Ferner *soll* die diskrete Ereignisanalyse vorzuziehen sein, wenn die Lokalisierung von Ereignissen in der Zeit sehr grob und ungenau ist, wenn also die zeitlichen Intervalle, in denen Ereignisse auftreten können, sehr breit sind (Yamaguchi 1991: 16; Allison 1982: 63).

Zeitdiskrete Ereignismodelle können nach Yamaguchi zeitkontinuierliche Modelle sehr gut approximieren, wenn die bedingten Wahrscheinlichkeiten, ein Ereignis zu einem diskreten Zeitpunkt zu haben, sehr klein sind (Yamaguchi 1991: 16f). Daraus läßt sich ableiten, daß die Zeitintervalle bei geringen Verweildauern sehr klein sein müssen, während sie bei großen Verweildauern recht groß sein dürfen (unter der Berücksichtigung der Verteilung der Episodendauern). Das Resultat ist dann meistens, wenn in den Rohdaten Informationen auf Monatsbasis vorliegen, ein Personen-Monat-Rekord-Datensatz, d.h. es gibt für jeden Monat und für jede Person einen eigenen Rekord. Damit wird zwar der Informationsgehalt der Daten voll ausgeschöpft, wie auch in Datensätzen für kontinuierliche Modelle, der Datensatz vergrößert sich jedoch immens und enthält zu einem sehr großen Teil redundante Informationen. Die Größe des nun zu verwendenden Datensatzes stellt zudem höhere Ansprüche an die Rechnerausstattung.

Zeitdiskreten Modellen liegt die Annahme zugrunde, daß die Rate innerhalb eines Zeitintervalls konstant ist. Man kann sich aber einen zugrunde liegenden kontinuierlichen Prozeß vorstellen, wenn die Abstände zwischen den diskreten Zeitpunkten sehr klein sind. Diese Annahme ist unproblematisch und eigentlich auch analog zu der Annahme in zeitkontinuierlichen Modellen, in denen Informationen verwendet werden, die auch nur zeitdiskret gemessen werden können. Diese Annahme basiert auf der Vorstellung, die Zeit sei hinreichend genau gemessen, um als adäquate Annäherung eines als kontinuierlich angenommenen Prozesses fungieren zu können.

Es stellt sich die Frage, ob diskrete Modelle tatsächlich die Möglichkeit bieten, mit einem Zeitaggregationsproblem umzugehen. Die Behauptung, daß die Verwendung diskreter Methoden zur Analyse von Ereignisdaten dann

angemessener sei, wenn die Zeit nur grob gemessen wurde, zeugt zumindest von einem gewissen Vertrauen in diese Art von ereignisanalytischen Modellen.

Zunächst einmal muß bei der Verwendung diskreter Modelle zur Analyse von Ereignisdaten die Annahme getroffen werden, daß die Rate zwischen zwei Zeitpunkten konstant ist. Ein geringer und somit akzeptabler Bias der Schätzungen ist nur zu erwarten, wenn der Abstand zwischen zwei Zeitpunkten relativ gering ist. Die Frage, wie gering dieser sein muß, läßt sich nicht ad hoc beantworten und aus den jeweils vorliegenden Daten auch nicht ableiten. Selbst wenn die Einteilung der Zeitachse als detailliert genug eingeschätzt wird, entfällt dadurch das Short-Spell-Problem nicht. Auch wird das Zeitaggregationsproblem nur im Zusammenhang mit der zeitlichen Skalierung des abhängigen Prozesses gesehen. Wenn der abhängige Prozeß zeitlich genau genug abgebildet ist, heißt dies nicht, daß dies auch für zeitveränderliche erklärende Variablen zutrifft. Für einen Prozeß mag die Erhebung auf Jahresbasis genau genug erscheinen. Für einen anderen Prozeß wäre eine solche zeitliche Basis aber zu grob.

Liegen für verschiedene Prozesse unterschiedliche Zeitbasen vor oder ist die zeitliche Messung eines Ereignisses zu ungenau, so besteht das Problem, nur unter zusätzlichen Annahmen und mit einer gewissen Ungenauigkeit eine zeitliche Ordnung zwischen den Prozessen herstellen zu können. Der kausale Zeitpfad ist damit „gestört“. Beeinflussen zwei Prozesse sich gegenseitig und kann keine exakte Ordnung hergestellt werden, so wird in Analysen nur noch ein Nettoeffekt auf den gegenseitigen Einfluß beider geschätzt. Die Komposition des Nettoeffekts hängt dabei nicht nur von der jeweiligen Stärke der Einflüsse ab, sondern auch von dem Anteil der durch die Zeitaggregation und durch Annahmen entstandenen falschen zeitlichen Ordnungen zwischen Prozessen bei einzelnen Analyseeinheiten. Auch nicht lineare Effekte können die Form ihres Einflusses verändern. Gleiches ließe sich auch bezüglich stetigen Modellen und Zeitaggregation anführen.

Es scheint also, als gäbe es keine Vorteile diskreter ereignisanalytischer Methoden gegenüber stetigen in bezug auf das Aggregationsproblem. Das Problem liegt weniger in den Analysemethoden als vielmehr in den Daten selbst.

4 Beispiele

Anhand von Daten der Lebensverlaufsstudie mit Informationen, die auf Monatsbasis vorliegen, soll die Adäquatheit diskreter Modelle im Hinblick auf das Problem der Zeitaggregation überprüft werden. Dazu werden die auf Monatsbasis vorliegenden Informationen sukzessive durch die Simulation von Beobachtungsintervallen zeitlich verschlechtert. Bisherige Simulationen dieser Art (vgl. zum Beispiel Bergström und Edin 1992) rundeten zur Über-

prüfung des Zeitaggregationsproblems die Episodendauern nur auf und ignorierten damit praktisch das Short-Spell-Problem.

4.1 Daten

Die Lebensverlaufsstudie entstammt dem Projekt *Lebensverläufe und sozialer Wandel*, welches unter der Leitung von Karl Ulrich Mayer am Max-Planck-Institut für Bildungsforschung in Berlin angesiedelt ist. Sie umfaßt mehrere Teilstudien, in denen zwischen den Jahren 1981 und 1992 retrospektive Lebensverläufe von Angehörigen unterschiedlicher Geburtskohorten in Ost- und Westdeutschland erhoben wurden. Vier Teilprojekte wurden in Westdeutschland und eines in Ostdeutschland durchgeführt.

Der den eigenen Analysen in dieser Arbeit zugrunde liegende Datensatz umfaßt die Geburtsjahrgänge 1929–1931, 1939–1941 und 1949–1951. In den Jahren 1981 bis 1983 wurden in dem Teilprojekt *Lebensverläufe und sozialer Wandel* Lebensverläufe von 2171 Mitgliedern der besagten Geburtsjahrgänge erhoben (Mayer und Brückner 1989). Die Geburtsjahrgänge 1919–1921, 1954–1956, 1959–1961 und die Ostdeutschen Geburtsjahrgänge fanden keinen Einzug in die vorliegenden Analysen. Die Beschränkung auf die Daten aus der ersten Erhebung hat keine theoretischen Gründe. Da hier nur anhand eines Beispieldatensatzes Berechnungen erfolgen sollen, wurde der Datensatz auf die drei besagten Geburtsjahrgänge beschränkt. Folgende Variablen wurden konstruiert:

Erwerbsstatus Es liegen Informationen über Beginn und Ende von Beschäftigungsverhältnissen und eine mögliche Zensierung zum Interviewzeitpunkt vor.

Prestige Das Berufsprestige für die jeweilige Beschäftigung liegt hier als Treimansscore vor (Treimann 1977, 1979), eine internationale Klassifikation des Berufsprestiges.

Kohortenzugehörigkeit Für die Kohortenzugehörigkeit (1929–1931, 1939–1941, 1949–1951) werden Dummyvariablen gebildet, wobei die Geburtskohorte 1929–1931 die Referenzkategorie darstellt.

Arbeitsmarkterfahrung Sie wird wie in Blossfeld und Rohwer (1995) gemessen durch die Zeit, die seit dem ersten Eintritt in den Arbeitsmarkt vergangen ist. Für die stetigen ereignisanalytischen Modelle wurden die Episoden zusätzlich auf Monatsbasis gesplittet, um eine möglichst große Vergleichbarkeit gegenüber den diskreten Modellen auf Einmonatsbasis zu erreichen.

Anzahl der bisherigen Beschäftigungen Sie wird gemessen durch die Nummer der derzeitigen Jobepisode minus 1. Es ist zu beachten, daß sich bei der Simulation von Beobachtungsintervallen die Anzahl der bisherigen Jobs ändert, wenn vorherige Episoden nicht beobachtet werden konnten.

Bildung Diese Variable ist der höchste Ausbildungsabschluß, welcher erhoben wurde. Es werden dabei Schuljahren bestimmte Abschlüssen zugewiesen, die zur Erreichung dieses Abschlusses nötig wären. Die Kodierung orientiert sich an Blossfeld (1985, 1992): kein Abschluß ist mit 8 Jahren, Hauptschulabschluß ohne Ausbildung mit 9 Jahren, Mittlere Reife ohne Ausbildung mit 10 Jahren, Hauptschulabschluß mit Ausbildung mit 11 Jahren, Realschulabschluß mit einer Ausbildung mit 12 Jahren, Abitur ohne Ausbildung mit 13 Jahren, Abitur mit Ausbildung mit 15 Jahren, Fachhochschulabschluß mit 17 Jahren und Universitätsabschluß oder höher ist mit 19 Jahren kodiert.

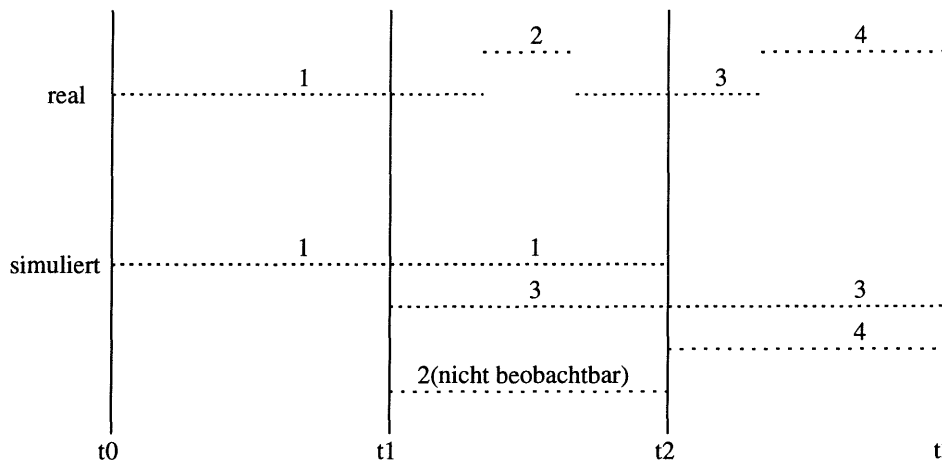
Verheiratet Diese Variable ist Dummykodiert und gibt an, ob eine Person verheiratet ist oder schon einmal verheiratet war. Sie ist damit zeitveränderlich. Zusätzlich wird in den Modellen noch eine Variable für verheiratete Männer mit aufgenommen.

4.2 Simulation von Beobachtungsintervallen

Für die vorliegenden Analysen wurden mit dem Programm *TACOS* (Steinhage 1999) Beobachtungsintervalle simuliert. Dadurch gehen zum einen die Episoden verloren, die zu kurz sind, als daß sie beobachtet werden können (Short-Spell-Problem). Zum anderen steigen die bedingten Wahrscheinlichkeiten, ein Ereignis zu einem diskreten Zeitpunkt zu haben, stark an (Zeitaggregationsproblem), wodurch die Güte der Daten sinkt. Ferner verändert sich auch die Verteilung der Episodendauern.

Der Algorithmus nach dem die Simulation vollzogen wird, kann folgendermaßen beschrieben werden: Es werden in gleichen Abständen Beobachtungen simuliert. Die erste Beobachtung könnte nur zeigen, in welchem Zustand sich die Untersuchungseinheiten befinden, bei Erwerbsepisoden also ob die Individuen erwerbstätig sind oder nicht. Erst bei einer zweiten Beobachtung wäre es möglich eine Veränderung auszumachen. Hat eine Person eine neue Beschäftigung angenommen, so wird die alte beendet worden sein. Die Simulation von Beobachtungsintervallen hat nun aber zur Folge, daß der genaue Zeitpunkt einer Veränderung nicht mehr ausgemacht werden kann (vereinfachend wird hier davon ausgegangen, daß die zugrunde liegenden Daten die *genauen* Zeitpunkte eines Wechsels angeben). Es kann nur noch angegeben werden, daß die Veränderung in dem Zeitintervall vor der Beobachtung, die die Veränderung ergeben hat, und nach der davor liegenden Beobachtung

Abbildung 1: Simulation von Beobachtungsintervallen



stattgefunden haben muß. Es kann in einem Zeitintervall nur eine Veränderung beobachtet werden. Das heißt, wenn mehr als einmal der Zustand (zum Beispiel die Beschäftigung) gewechselt wird, läßt sich dies nicht mehr beobachten; mit anderen Worten: Wird ein neuer Zustand nach einer Beobachtung eingenommen und wird dieser Zustand vor der nächsten Beobachtung wieder aufgegeben, so kann dieser Wechsel nicht beobachtet werden. Je kürzer die Verweildauern in einem Zustand sind beziehungsweise je größer das Zeitintervall zwischen zwei Beobachtungen ist, desto höher ist die Wahrscheinlichkeit, einzelne Episoden nicht beobachten zu können.

Die Prozedur der Simulation von Beobachtungsintervallen, wie sie in dem Programm *TACOS* implementiert ist, kann an der Abbildung 1 veranschaulicht werden. Gegeben seien vier reale Episoden (1, 2, 3 und 4). Wird eine Simulation durchgeführt mit einem Simulationsintervall der Länge $t_1 - t_0$, so wird die Episode 1 gesplittet. Der erste Split reicht von t_0 bis t_1 und ist rechtsenziert, der zweite reicht von t_1 bis t_2 und wird durch ein Ereignis beendet. Die Episode 2 konnte nicht beobachtet werden, da sie erst nach dem Zeitpunkt t_1 beginnt und noch vor dem Zeitpunkt t_2 endet. Die Episode 3 wurde gesplittet und reicht insgesamt von t_1 bis t_3 und wird durch ein Ereignis beendet. Die Episode 4 liegt in dem Intervall zwischen t_2 und t_3 .

Die Simulation von Beobachtungsintervallen ist auch in der Hinsicht interessant, als daß mit zunehmender Größe der Intervalle eine genaue Spezifizierung von Ursache und Wirkung nicht mehr möglich ist. Wenn in einem Zeitintervall beispielsweise eine Frau, welche in einer nichtehelichen Lebensgemeinschaft lebt, schwanger wird und sie und ihr Partner infolge dessen heiraten, so kann die zeitliche Abfolge beider Ereignisse nicht mehr ausgemacht werden (Kelly und McGrath 1988). Es ergeben sich damit ähnliche

Probleme wie bei einem Querschnitt (vgl. Steinhage und Blossfeld 1999). Zum anderen besteht die Möglichkeit, wenn sich ein paralleler Prozeß dynamisch über die Zeit hinweg verändert (vgl. Kelly und McGrath 1988: 27; Blossfeld und Rohwer 1995: 15), daß diese Dynamik durch die Größe des gewählten Intervalls unbeobachtet bleibt. Wäre das gewählte Beobachtungsintervall größer als die Zeit, die ein paralleler Prozeß braucht, sich auf den abhängigen Prozeß auszuwirken, und ist die Veränderung der Wirkungsweise dieses parallelen Prozesses zeitlich begrenzt auf eine Dauer, die kleiner ist als das gewählte Intervall, so wird nur der durchschnittliche Einfluß des parallelen Prozesses (nicht seine Dynamik) auf den abhängigen Prozeß geschätzt (vgl. hierzu Blossfeld, Manting, Rohwer 1993). Es wäre sogar möglich, daß ein paralleler Prozeß aufgrund des Beobachtungsintervalls keinen Einfluß auf den abhängigen Prozeß zu haben scheint.

4.3 Exponentialmodelle

Nachdem die Simulation von Beobachtungsintervallen verdeutlicht worden ist und die Verwendung findenden Daten beschrieben worden sind, werden einige Beispielmmodelle miteinander verglichen. Es wurde ein stetiges Modell und ferner diskrete Modelle mit Beobachtungsintervallen von einem, 12, 24, 48 und 96 Monaten geschätzt. Konnte ein Ereignis innerhalb eines Beobachtungsintervalls festgestellt werden, so wurde das Ende des jeweiligen Intervalls als Zeitpunkt des Ereigniseintritts angenommen. Die zeitliche Ordnung zwischen Variablen ist in allen Modellen kontrolliert worden, um Veränderungen in den geschätzten Parametern ausschließlich auf das Zeitaggregations- und das Short-Spell-Problem zurückführen zu können.

Bei Beschäftigungsepisoden mag es zunächst überflüssig erscheinen, solch große Beobachtungsintervalle zu verwenden. Man kann sich jedoch einen anderen Prozeß vorstellen, der nur auf Jahresbasis erhoben wurde, aber innerhalb eines Jahres eine starke Dynamik aufweist. Von daher können die Ergebnisse bezüglich der recht großen Beobachtungsintervalle bei den Beschäftigungsepisoden auf andere, schneller ablaufende Prozesse mit kleineren Beobachtungsintervallen übertragen werden.

Das einfachste Modell ist ein Exponentialmodell, welches eine zeitkonstante Rate zur Annahme hat. Folgende Modelle wurde konstruiert:

$$r(t|X_i, X_k(t)) = \exp(\alpha + \sum_i \beta_i X_i + \sum_k \beta_k X_k(t)) \quad (5)$$

für das stetige Modell und

$$\lambda(t|X_i, X_k(t)) = \frac{\exp(\alpha + \sum_i \beta_i X_i + \sum_k \beta_k X_k(t))}{1 + \exp(\alpha + \sum_i \beta_i X_i + \sum_k \beta_k X_k(t))} \quad (6)$$

für die diskreten Modelle. Es soll die Rate geschätzt werden, eine Beschäftigung zu verlassen, unabhängig davon, ob das Resultat eine andere Beschäftigung ist oder nicht.

Wie aus Tabelle 1 ersichtlich ist, nimmt die Anzahl der beobachteten Ereignisse mit zunehmender Größe der Beobachtungsintervalle ab. Bei den Modellen auf Einmonatsbasis (keine Beobachtungsintervalle wurden simuliert) liegen 4126 Ereignisse, bei einer Breite der Beobachtungsintervalle von 12 Monaten 3651, bei 24 Monaten 3132, bei 48 Monaten 2327 und bei 96 Monaten 1418 Ereignisse vor. Zwischen diskretem und stetigem Modell auf Einmonatsbasis besteht bezüglich der Ereignisanzahl logischer Weise kein Unterschied. Mit zunehmender Größe der Beobachtungsintervalle steigt die Wahrscheinlichkeit, nur noch verhältnismäßig lange Episoden zu beobachten. Ferner wird die Episodenlänge zunehmend überschätzt, da nicht mehr mit Sicherheit gesagt werden kann, wann eine Episode beendet wurde. Es ist nur bekannt ab wann, wenn sie beendet wurde, sie mit Sicherheit nicht mehr andauerte. Die genaue Lokalisierung des Ereignisses im Zeitfenster zwischen zwei Beobachtungen ist damit nicht mehr möglich. Bei einer Teilung dieses Zeitfensters in zwei Teile wird mit zunehmender Größe der Beobachtungsintervalle die Wahrscheinlichkeit größer, daß das Ende der Episode, welche in dem Zeitfenster ein Ereignis gehabt haben muß, in der ersten Fensterhälfte liegt. Das Resultat ist ein Zeitaggregationsbias (vgl. Petersen 1995, 1993: 435ff).

Bis auf leichte Unterschiede in den Nachkommastellen der Parameter für das diskrete und das stetige Modell gibt es zwischen beiden keine Unterschiede. Je mehr Arbeitsmarkterfahrung eine Person hat, desto geringer ist ihre Neigung, den Job zu verlassen (vgl. Becker 1975). Die Anzahl der bisherigen Beschäftigungen erhöht dagegen das Risiko, die aktuelle Beschäftigung zu verlassen.² Gleiches gilt für den Effekt von Bildung. Je höher die Bildung, desto geringer die Verweildauer in der Beschäftigung. Je höher das Prestige der aktuellen Beschäftigung ist, desto geringer die Neigung, sie zu verlassen, da es mit zunehmendem Prestige unwahrscheinlicher wird, eine Beschäftigung mit einem noch höheren Prestige zu bekommen beziehungsweise desto weniger attraktiv wird es, die aktuelle Beschäftigung zu verlassen. Die Kohorte 1939–1941 hat keine signifikant höhere Rate die Beschäftigung zu verlassen, als die Referenzkohorte 1929–1931. Die Geburtskohorte 1949–1951 hat dagegen eine signifikant höhere Rate. Verheiratete Frauen verlassen schneller eine Beschäftigung als verheiratete Männer und unverheiratete Personen. Verhei-

²Nach der Segmentationstheorie kann dieser Effekt auf das geringere betriebsspezifische Humankapital zurückgeführt werden, da für Personen mit vielen bisherigen Beschäftigungsverhältnissen die Wahrscheinlichkeit größer ist, jeweils nur relativ kurz in einer Beschäftigung verweilt zu haben. Diese Personen würden daher eher auf dem sekundären Arbeitsmarkt anzutreffen sein (vgl. Döringer und Piore 1971; Piore 1969).

Tabelle 1: Stetiges Exponentialmodell und diskrete Exponentialmodelle mit Simulationsintervallen von einem, 12, 24, 48 und 96 Monaten

	stetig	diskret 1	diskret 12	diskret 24	diskret 48	diskret 96
Konstante	-3,5454** (-41,6739)	-3,5191** (-40,9831)	-1,4312** (-14,6050)	-1,1123** (-9,9892)	-1,1475** (-8,2297)	-1,2573** (-6,2865)
Arbeitsmarkterfahrung	-0,0072** (-25,5224)	-0,0073** (-25,6535)	-0,0055** (-17,8659)	-0,0033** (-10,0732)	0,0009** (2,4599)	0,0080** (15,6461)
Anzahl der bisherigen Jobs	0,1387** (14,1984)	0,1409** (14,2365)	0,0469** (2,8908)	-0,1094** (-4,8278)	-0,5026** (-13,7516)	-1,5560** (-20,3147)
Bildung	0,0333** (4,7506)	0,0340** (4,7958)	0,0352** (4,3462)	0,0377** (4,0868)	0,0387** (3,3724)	0,0163 (1,0004)
Prestige	-0,0163** (-9,3744)	-0,0166** (-9,4593)	-0,0163** (-8,1589)	-0,0168** (-7,4597)	-0,0170** (-6,1625)	-0,0161** (-4,2975)
Kohorte 1939–1941	0,0642 (1,6659)	0,0649 (1,6703)	0,1041* (2,3716)	0,1491** (3,0375)	0,2480** (4,1611)	0,1458 (1,7840)
Kohorte 1949–1951	0,1304** (3,1153)	0,1327** (3,1400)	0,1512** (3,0613)	0,1683** (2,9521)	0,2129** (2,9365)	0,2906** (2,8546)
Verheiratet	0,4107** (9,3873)	0,4176** (9,4348)	0,7119** (14,2383)	0,8678** (15,1976)	1,0971** (15,1693)	1,2634** (11,6829)
Verheiratete Männer	-0,7216** (-15,4316)	-0,7329** (-15,5521)	-0,7791** (-15,5495)	-0,7759** (-14,1396)	-0,6921** (-10,7089)	-0,6078** (-7,1737)
Log-Likelihood	-21276	-21238	-1660	-7711	-4925	-2583
Subepisoden	331362	331362	30818	16749	9171	5095
Ereignisse	4126	4126	3651	3132	2327	1418

* $p \leq 0.05$, ** $p \leq 0.01$; Coeff./std. error in Klammern

Daten: Lebensverlaufsstudie; eigene Berechnungen

ratete Männer³ haben dagegen eine geringere Neigung, eine Beschäftigung zu verlassen. Nach der Familienökonomie (Becker 1993) erhöht sich für verheiratete Frauen die Neigung, die Erwerbstätigkeit zu verlassen, um sich stärker im Haushalt und bei der Kindererziehung zu engagieren. Dagegen werden verheiratete Männer in Anbetracht gestiegener ökonomischer Verantwortung eher erwerbstätig sein.

Bei der Betrachtung der Ergebnisse für die diskreten Modelle nach zunehmender Größe der Beobachtungsintervalle sind einige Auffälligkeiten zu verzeichnen. Ab einem Beobachtungsintervall von 48 Monaten ist eine Verkehrung des Vorzeichens für den Parameter der Arbeitsmarkterfahrung bei weiterhin hoher Signifikanz zu beobachten. Eine hohe Arbeitsmarkterfahrung scheint nicht mehr zu einer Verstetigung des Beschäftigungsverhältnisses zu führen, sondern umgekehrt dazu, daß die Beschäftigung schneller verlassen wird.

Bei dem Parameter für die Anzahl der bisherigen Beschäftigungen ist eine Verkehrung des Vorzeichens ab einem Beobachtungsintervall von 24 Monaten zu verzeichnen. Mit zunehmender Größe des sich im negativen Bereich befindenden Parameters erhöht sich auch seine Signifikanz. Die Verkehrung des Vorzeichens kann darauf zurückgeführt werden, daß es mit zunehmender Größe der Beobachtungsintervalle zu einer Selektion kommt. Es steigt die Wahrscheinlichkeit, relativ kurze Episoden nicht mehr beobachten zu können. Das steigende Übergewicht spezifischer Jobepisoden, nämlich besonders stabiler, führt letztendlich zu der besagten Parameterveränderung.

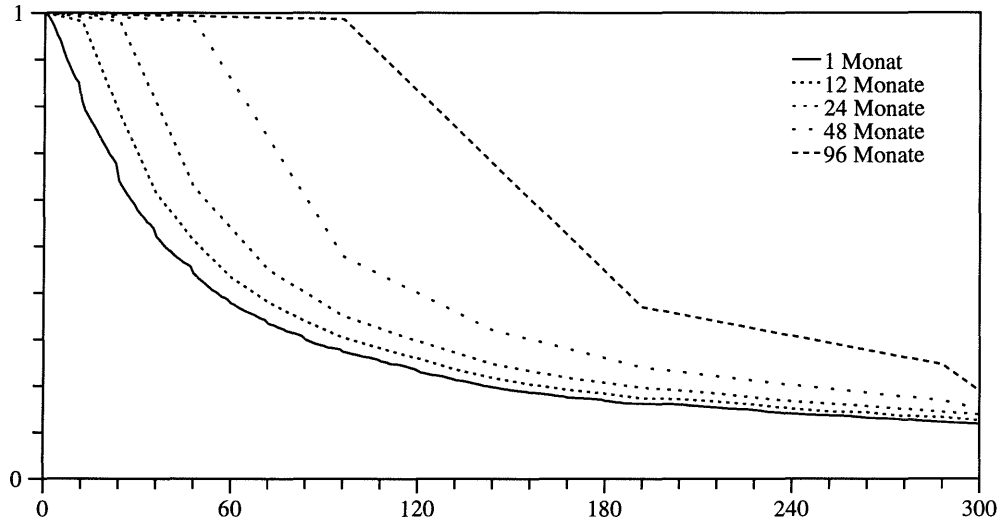
Der Parameter für die Bildung bleibt relativ konstant, verliert bei einem Beobachtungsintervall von 96 Monaten aber seine Signifikanz. Das Prestige scheint dagegen nicht von dieser Simulation beeinflusst zu werden.

Die sich zunächst nicht signifikant von der Referenzkohorte unterscheidende Geburtskohorte 1939–1941 zeigt bei Beobachtungsintervallen von 12, 24 und 48 Monaten signifikante Unterschiede. Dieses kann als ein Indiz dafür angesehen werden, daß ein Zusammenhang zwischen der Kohortenzugehörigkeit und einer weiteren Variable besteht, die durch die Simulation von Beobachtungsintervallen beeinflusst wird. Mit zunehmender Größe der Beobachtungsintervalle unterscheidet sich die Geburtskohorte 1949–1951 immer stärker von der Referenzkohorte.

Der Parameter für verheiratete Frauen ist über die Beobachtungsintervalle ebenfalls ansteigend, positiv und hoch signifikant. Der Parameter für verheiratete Männer befindet sich zwar immer im negativen Bereich und ist hoch signifikant, er ist aber für die besagte Gruppe nicht ohne den Parameter für *Verheiratet* zu interpretieren. Werden beide Parameter zusammen betrachtet, so ist zu erkennen, daß verheiratete Männer ab einem Beobachtungsintervall

³Es ist zu beachten, daß für verheiratete Männer zwei Parameter von Bedeutung sind: *Verheiratet* und *Verheiratete Männer*.

Abbildung 2: Survivor-Funktionen für verschiedene Zeitaggregationen



von 24 Monaten eine zunehmend höhere Rate haben, ihre Beschäftigung zu verlassen, als unverheiratete Personen.

In Abbildung 2 sind Survivor-Funktionen für verschiedene Zeitaggregationen dargestellt. Die Survivor-Funktionen wurden mit der Produktlimit-schätzung von Kaplan und Meier (1958) berechnet. Es seien q Zeitpunkte gegeben, zu denen mindestens ein Ereignis stattfindet:

$$\tau_1 < \tau_2 < \tau_3 < \dots < \tau_q. \quad (7)$$

Sind für alle $l = 1, \dots, q$ mit der Konvention, daß $\tau_0 = 0$, die Quantitäten E_l (Anzahl der Ereignisse zu τ_l), Z_l (Anzahl der zensierten Episoden, die in dem Intervall τ_{l-1} und τ_l enden) und R_l (die Anzahl der Episoden, die noch dem Risiko eines Ereignisses ausgesetzt sind: $Z_l + E_l$) bestimmt, so kann die Survivor-Funktion $S(t)$ auf folgende Art berechnet werden:

$$S(t) = \prod_{l: \tau_l < t} \left(1 - \frac{E_l}{R_l}\right). \quad (8)$$

Mit zunehmender Größe der Beobachtungsintervalle steigt die Verweildauer der Episoden an. Dieses resultiert aus zwei Sachverhalten. Zum einen wird die Dauer der Episoden zunehmend überschätzt. Zum anderen steigt die Wahrscheinlichkeit, sehr kurze Episoden nicht mehr beobachten zu können. Weiter ist zu ersehen, daß auf Einmonatsbasis sich die Mehrzahl der Zustandswechsel zwischen 1 und 96 Monaten ereignet, während dies auf 96-Monatsbasis (simulationsbedingt) zwischen 97 und 192 Monaten der Fall ist. Wie aus

der Abbildung ferner ersichtlich ist, werden die Episodendauern zunehmend länger⁴ und die Varianz in den Episodendauern sinkt.

Anhand von Survivor-Funktionen oder ihren Äquivalenten lassen sich auch Verteilungen der bedingten Wahrscheinlichkeiten, in einem Intervall oder zu einem diskreten Zeitpunkt ein Ereignis zu haben, visualisieren. Wie an Abbildung 2 gesehen werden kann, ist die von Yamaguchi (1991: 16f) genannte Grenze für die bedingten Wahrscheinlichkeiten von <0.1 schon ab einer Größe der Beobachtungsintervalle von 12 Monaten überschritten.

Bei einem einfachen Exponentialmodell bestehen zwischen diskreter und stetiger Modellierung auf Einmonatsbasis keine gravierenden Unterschiede. Wenn davon ausgegangen wird, daß die Modelle auf Einmonatsbasis vollständigere Information enthalten als die anderen Modelle, so treten schon kleinere Abweichungen bei einer Modellierung auf Jahresbasis auf. Modelle mit einer noch groberen Einteilung der Zeitachse führen bei einigen Variablen zu gravierenden Veränderungen in den geschätzten Parametern, was zu Fehlinterpretationen führen kann.

4.4 Periodenspezifische Exponentialmodelle

Die Veränderungen in den Ergebnissen des einfachen Exponentialmodells über die unterschiedlichen Zeitaggregationen könnten auch darauf zurückzuführen sein, daß die Zeitabhängigkeit der Rate nicht kontrolliert wurde. Daher werden im folgenden periodenspezifische Exponentialmodelle für die gleichen aus der Simulation hervorgegangenen Daten geschätzt. Die Modelle sind wie folgt konstruiert:

$$r(t|X_i, X_k(t)) = \exp\left(\sum_{l=1}^{L^*} \alpha_l d_l + \sum_i \beta_i X_i + \sum_k \beta_k X_k(t)\right) \quad (9)$$

für das stetige Modell und

$$\lambda(t|X_i, X_k(t)) = \frac{\exp(\alpha_0 + \sum_{l=1}^{L^*-1} \alpha_l d_l + \sum_i \beta_i X_i + \sum_k \beta_k X_k(t))}{1 + \exp(\alpha_0 + \sum_{l=1}^{L^*-1} \alpha_l d_l + \sum_i \beta_i X_i + \sum_k \beta_k X_k(t))} \quad (10)$$

für die diskreten Modelle. In einem periodenspezifischen Exponentialmodell wird angenommen, daß die Rate innerhalb jeder Periode d_l (mit $l = 1, 2, \dots, L^*$) konstant ist, zwischen den einzelnen Perioden aber variieren kann. Die Ergebnisse sind den Tabellen 2 und 3 zu entnehmen.

Die Parameter, bis auf die für die Zeitperioden, sind bei dem stetigen und dem diskreten Modell auf Einmonatsbasis fast identisch. Selbst beim

⁴Dies wäre selbst der Fall, wenn nicht t , sondern $t - 0.5$ als Ende der Episode angenommen worden wäre.

Vergleich mit den entsprechenden Ergebnissen des einfachen Exponentialmodells sind keine gravierenden Unterschiede erkennbar. Im stetigen Modell ist tendenziell ein Anstieg und nach 60 Monaten ein Abfallen der Rate über die Zeitperioden zu beobachten. Das entsprechende diskrete Modell auf Einmonatsbasis, bei dem die ersten 12 Monate die Referenzkategorie darstellen, zeigt nur in den Perioden 12 bis 24 und 24 bis 36 Monate eine im Vergleich zur Referenzkategorie höhere Rate. Dieses kann auf die Wahl der Referenzkategorie zurückgeführt werden.

Trotz der unterschiedlichen Skalierung der Zeit können für das diskrete Modell mit Beobachtungsintervallen von 12 Monaten gleiche Perioden gebildet werden. Es unterscheiden sich alle Perioden signifikant von der Referenzperiode (1. Jahr). Hier ist ein ähnlicher Verlauf beobachtbar wie im stetigen Modell. Die Rate steigt tendenziell bis zu einer Dauer von 2 bis 3 Jahren an und fällt dann ab. Je länger eine Person nach dem Maximum der Rate in der aktuellen Beschäftigung verweilt, desto unwahrscheinlicher ist es, daß sie diese verläßt. Die abfallende Rate, was als ein Effekt der Akkumulation von betriebsspezifischem Humankapital interpretiert werden kann, ist durch andere Untersuchungen gut dokumentiert (vgl. DiPrete 1981; Kandel und Yamaguchi 1987; Sørensen und Tuma 1981; Tuma 1976). Bis auf den Parameter für die bisherige Anzahl der Beschäftigungen vergrößern die Parameter ihren Abstand gegenüber Null.

Erst wenn die Beobachtungsintervalle weiter vergrößert werden, zeigen sich deutliche Unterschiede in den Ergebnissen. Zunächst ist zu sagen, daß die Perioden immer breiter werden und schlechter mit denen der vorhergehenden Modelle zu vergleichen sind, da durch die immer ungenauere zeitliche Skalierung Perioden zusammengefaßt werden müssen. Bei den Modellen mit Beobachtungsintervallen von 24, 48 und 96 Monaten ist eine abfallende Rate zu beobachten. Der Anstieg in der Rate zu Beginn, der noch in den vorherigen Modellen sichtbar war, geht durch die Zeitaggregation ganz verloren.

Der Parameter für den Effekt der Arbeitsmarkterfahrung ist durchgehend signifikant und negativ. Entgegen dem einfachen Exponentialmodell verändert der Koeffizient für die Variable der Anzahl der bisherigen Beschäftigungen ab einem Beobachtungsintervall von 24 Monaten nicht sein Vorzeichen. Er verliert aber ab 48 Monaten seine Signifikanz. Selbst bei einer weiteren Zeitaggregation verändert er sein Vorzeichen nicht. Dieses bestätigt die Erklärung der Vorzeichenwechsel der betreffenden Parameter im einfachen Exponentialmodell durch eine Selektion längerer Episoden. Dieser Selektionseffekt konnte durch die Berücksichtigung einer Zeitabhängigkeit der Rate zum Teil kontrolliert werden, so daß der Parameter für die Anzahl der bisherigen Beschäftigungen nicht sein Vorzeichen verändert. Dieses läßt darauf schließen, daß insbesondere zeitveränderliche Variablen bei Zeitaggregation in einem fehlspezifizierten Modell leichter zu problematischen Para-

Tabelle 2: Periodenspezifisches Exponentialmodell für ein stetiges Modell und für Simulationsintervalle von einem und 12 Monaten

	stetig	diskret 1M	diskret 12M
Konstante	—	-3,5578**	-3,7343**
	—	(-40,0027)	(-25,7357)
1–12 Monate	-3,6784**	Referenz	Referenz
	(-41,5002)	—	—
12–24 Monate	-3,4309**	0,0961*	2,7037**
	(-38,1969)	(2,0714)	(23,8019)
24–36 Monate	-3,4196**	0,1080*	2,9452**
	(-36,8134)	(2,0568)	(25,5356)
36–48 Monate	-3,5183**	0,0196	2,7128**
	(-35,7522)	(0,3188)	(22,5147)
48–60 Monate	-3,4578**	0,0365	2,7313**
	(-33,5249)	(0,5253)	(21,8425)
60–72 Monate	-3,6026**	-0,0586	2,5828**
	(-32,1404)	(-0,7226)	(19,4581)
72–84 Monate	-3,6000**	-0,0621	2,5436**
	(-30,1675)	(-0,6831)	(18,1833)
84–96 Monate	-3,6850**	-0,2121	2,5737**
	(-28,0084)	(-1,9433)	(17,4962)
≥96 Jahre	-3,5274**	-0,0327	2,5783**
	(-32,3323)	(-0,4253)	(19,2534)
Arbeitsmarkterfahrung	-0,0073**	-0,0070**	-0,0075**
	(-19,9818)	(-19,1404)	(-16,8277)
Anzahl der bisherigen Jobs	0,1419**	0,1329**	0,1260**
	(12,6839)	(11,6562)	(6,3628)
Bildung	0,0339**	0,0336**	0,0401**
	(4,8249)	(4,7462)	(4,8079)
Prestige	-0,0165**	-0,0167**	-0,0173**
	(-9,4754)	(-9,5145)	(-8,4228)
Kohorte 1939–1941	0,0638	0,0669	0,0808
	(1,6546)	(1,7210)	(1,8065)
Kohorte 1949–1951	0,1313**	0,1351**	0,1383**
	(3,1368)	(3,1937)	(2,7500)
Verheiratet	0,4049**	0,4196**	0,6257**
	(9,2352)	(9,4432)	(12,2440)
Verheiratete Männer	-0,7204**	-0,7252**	-0,8039**
	(-15,3393)	(-15,3226)	(-15,6490)
Log-Likelihood	-21253	-21230	-9949
Subepisoden	331362	331362	30818
Ereignisse	4126	4126	3615

* $p \leq 0.05$, ** $p \leq 0.01$; Coeff./std. error in Klammern

Daten: Lebensverlaufsstudie; eigene Berechnungen

Tabelle 3: Periodenspezifisches Exponentialmodell für Simulationsintervalle von 24, 48 und 96 Monaten

	diskret 24M		diskret 48M		diskret 96M
Konstante	-3,8575**		3,9671**		-4,5633**
	(-23,3884)		(-19,2781)		(-14,8684)
0–2 Jahre	Referenz	0–4	Referenz	0–8	Referenz
	—		—		—
2–4 Jahre	3,6851**	4–8	4,4465**	>8	5,5957**
	(30,2950)		(32,3559)		(27,8758)
4–6 Jahre	3,5547**	>8	4,0612**		—
	(27,7365)		(25,4824)		—
6–8 Jahre	3,4338**		—		—
	(24,9295)		—		—
>8 Jahre	3,3275**		—		—
	(22,9002)		—		—
Arbeitsmarkterfahrung	-0,0074**		-0,0068**		-0,0079**
	(-14,2924)		(-10,9738)		(-10,4275)
Anzahl der bisherigen Jobs	0,0982**		0,0291		0,0185
	(3,3536)		(0,6019)		(0,1994)
Bildung	0,0483**		0,0532**		0,0602**
	(4,8049)		(3,9000)		(2,8933)
Prestige	-0,0182**		-0,0181**		-0,0191**
	(-7,5117)		(-5,6216)		(-4,0262)
Kohorte 1939–1941	0,0911		0,1110		0,0376
	(1,7547)		(1,6567)		(0,3814)
Kohorte 1949–1951	0,1381*		0,1577		0,1499
	(2,2682)		(1,8308)		(1,0330)
Verheiratet	0,6741**		0,7247**		0,9431**
	(10,9674)		(8,3270)		(6,2400)
Verheiratete Männer	-0,8542**		-0,8557**		-0,9596**
	(-14,5514)		(-11,3485)		(-8,4707)
Log-Likelihood	-6555		-3645		-1656
Subepisoden	16749		9171		5095
Ereignisse	3132		2327		1418

* $p \leq 0.05$, ** $p \leq 0.01$; Coeff./std. error in Klammern

Daten: Lebensverlaufsstudie; eigene Berechnungen

meterschätzungen führen und dies vor allem dann, wenn es wie hier zu einer Selektion von Episoden mit einer längeren Verweildauer kommt.

Die Kohorte 1949–1951 unterscheidet sich ab einem Beobachtungsintervall von 48 Monaten nicht mehr signifikant von der Referenzkohorte. Ferner verringert sich der Unterschied zwischen verheirateten Männern und verheirateten Frauen zusehends. Im Gegensatz zum einfachen Exponentialmodell haben verheiratete Frauen durchgehend eine größere Neigung als unverheiratete Personen und unverheiratete Personen eine größere Neigung als verheiratete Männer ihr Beschäftigungsverhältnis zu beenden. Der Trend über die

Simulationsintervalle hinweg geht dahingehend, daß sich verheiratete Männer immer weniger von unverheirateten Personen unterscheiden.

Zusammenfassend ist zu sagen, daß die Kontrolle der Zeitabhängigkeit der Rate die Ergebnisse in so weit beständiger machte, als daß die Einflußrichtung der Kovariaten auf den abhängigen Prozeß bis zu einem Beobachtungsintervall von 24 Monaten recht stabil blieben. Aufgrund der Zeitaggregation und des Short-Spell-Problems änderte sich jedoch der Zeitverlauf der Rate, so daß hier zumindest Vorsicht bei ihrer Interpretation geboten sein sollte.

4.5 Competing-Risk Exponentialmodelle

Bisher ist nur die Heterogenität in bezug auf einige Variablen und eine Zeitabhängigkeit der Rate kontrolliert worden. Es wurde davon abgesehen, in welchen Zustand die Individuen wechseln, wenn sie ihren Job verlassen. Geschätzt werden soll nun ein Modell konkurrierender Risiken.

$$r_d(t|X_i, X_k(t)) = \exp(\alpha_d + \sum_i \beta_{di} X_{di} + \sum_k \beta_{dk} X_{dk}(t)) \quad (11)$$

für das stetige Modell und

$$\lambda_d(t|X_i, X_k(t)) = \frac{\exp(\alpha_d + \sum_i \beta_{di} X_{di} + \sum_k \beta_{dk} X_{dk}(t))}{1 + \exp(\alpha_d + \sum_i \beta_{di} X_{di} + \sum_k \beta_{dk} X_{dk}(t))} \quad (12)$$

für die diskreten Modelle, mit $d = 1, 2, 3$, wobei 1 einen Abstieg, 2 einen vom Prestige her ähnlichen Job und 3 einen beruflichen Aufstieg, der definiert ist als Vergrößerung des Prestiges um 20% gegenüber dem vorherigen Job, meint.⁵ Die Rate, eine Beschäftigung zu verlassen und in eine der drei Richtungen mobil zu sein, ist:

$$r = \sum_{d \in D_j} r_d. \quad (13)$$

D_j kennzeichnet alle möglichen Zielzustände, die vom Ausgangszustand aus erreicht werden können. Wegen des Umfangs der Tabellen sollen die Ergebnisse nur grob dargestellt werden.

Ein Vergleich der Modelle für stetige und diskrete Ereignisanalyse basierend auf Einmonatsbasis (Tabellen 4 und 5) zeigt, daß der einzige Unterschied, abgesehen von geringfügigen Unterschieden in den Parametern, darin besteht, daß der Parameter für Prestige im diskreten Modell für horizontale Mobilität nun auf dem 5% Niveau signifikant ist.

⁵Im Falle von Erwerbslosigkeitsphasen wird von keiner Veränderung des Prestiges ausgegangen.

Tabelle 4: Competing-Risk Exponentialmodell für ein stetiges Modell mit Splits auf Monatsbasis

	aufwärts	horizontal	abwärts
Konstante	-3,5374** (-15,3404)	-4,6013** (-41,3392)	-4,6926** (-27,8887)
Arbeitsmarkterf.	-0,0064** (-8,9018)	-0,0078** (-19,7392)	-0,0064** (-12,9215)
Anzahl bish. Jobs	0,1422** (6,2863)	0,1573** (12,1649)	0,0876** (4,4571)
Bildung	0,1922** (11,4105)	0,0351** (3,8385)	-0,0824** (-5,8787)
Prestige	-0,1249** (-24,8478)	-0,0043 (-1,8612)	0,0120** (3,9135)
Kohorte 1939–1941	0,1309 (1,3916)	0,1315* (2,5188)	-0,0112 (-0,1538)
Kohorte 1949–1951	0,2577* (2,4912)	0,1652** (2,8965)	0,0756 (0,9793)
Verheiratet	0,0419 (0,3315)	0,0269 (0,4161)	1,0100** (13,8300)
Verheir. Männer	-0,5094** (-3,8334)	-0,2770** (-4,0898)	-1,3427** (-16,4290)
Log-Likelihood	-24874	-24874	-24874
Splits	327900	329496	328438
Ereignisse	664	2260	1202

* $p \leq 0.05$, ** $p \leq 0.01$; Coeff./std. error in Klammern

Quelle: Lebensverlaufsstudie; eigene Berechnungen

Tabelle 5: Competing-Risk Exponentialmodell für Simulationsintervalle von einem Monat

	aufwärts	horizontal	abwärts
Konstante	-3,5137** (-15,1798)	-4,5750** (-40,8832)	-4,6656** (-27,6514)
Arbeitsmarkterf.	-0,0065** (-9,0248)	-0,0079** (-19,8791)	-0,0065** (-13,0660)
Anzahl bish. Jobs	0,1456** (6,4047)	0,1593** (12,2266)	0,0897** (4,5470)
Bildung	0,1946** (11,4313)	0,0359** (3,8962)	-0,0821** (-5,8347)
Prestige	-0,1258** (-24,8091)	-0,0046* (-1,9933)	0,0118** (3,8234)
Kohorte 1939–1941	0,1320 (1,3985)	0,1324* (2,5255)	-0,0102 (-0,1397)
Kohorte 1949–1951	0,2625* (2,5253)	0,1676** (2,9249)	0,0783 (1,0108)
Verheiratet	0,0499 (0,3933)	0,0335 (0,5148)	1,0171** (13,8657)
Verheir. Männer	-0,5248** (-3,9395)	-0,2883** (-4,2399)	-1,3530** (-16,5218)
Log-Likelihood	-4257	-12999	-7557
Splits	327900	329496	328438
Ereignisse	664	2260	1202

* $p \leq 0.05$, ** $p \leq 0.01$; Coeff./std. error in Klammern

Quelle: Lebensverlaufsstudie; eigene Berechnungen

Tabelle 6: Competing-Risk Exponentialmodell für Simulationsintervalle von 12 Monaten

	aufwärts	horizontal	abwärts
Konstante	-1,2502** (-4,9809)	-2,5663** (-20,6822)	-2,5504** (-13,9192)
Arbeitsmarkterf.	-0,0050** (-6,7997)	-0,0055** (-13,2568)	-0,0055** (-10,2135)
Anzahl bish. Jobs	0,0818* (2,1477)	0,0569** (2,6766)	-0,0118 (-0,3948)
Bildung	0,1944** (10,0955)	0,0377** (3,6754)	-0,0805** (-5,2355)
Prestige	-0,1284** (-22,8861)	-0,0032 (-1,2453)	0,0107** (3,1700)
Kohorte 1939–1941	0,1942 (1,9094)	0,1867** (3,2282)	0,0018 (0,0228)
Kohorte 1949–1951	0,2874* (2,4795)	0,2156** (3,3043)	0,0583 (0,6800)
Verheiratet	0,3815** (2,9584)	0,2667** (3,8126)	1,4501** (17,7834)
Verheir. Männer	-0,6454** (-4,9384)	-0,3136** (-4,4774)	-1,4160** (-16,9426)
Log-Likelihood	-2490	-6969	-4209
Splits	27768	29141	28243
Ereignisse	601	1974	1076

* $p \leq 0.05$, ** $p \leq 0.01$; Coeff./std. error in Klammern

Quelle: Lebensverlaufsstudie; eigene Berechnungen

Tabelle 7: Competing-Risk Exponentialmodell für Simulationsintervalle von 24 Monaten

	aufwärts	horizontal	abwärts
Konstante	-1,0280** (-3,7124)	-2,1920** (-15,8228)	-2,2633** (-11,2437)
Arbeitsmarkterf.	-0,0025** (-3,3035)	-0,0031** (-7,2127)	-0,0037** (-6,4966)
Anzahl bish. Jobs	-0,0763 (-1,4668)	-0,1124** (-3,8255)	-0,1706** (-4,2070)
Bildung	0,2058** (9,5814)	0,0339** (2,9369)	-0,0686** (-4,0335)
Prestige	-0,1301** (-20,7246)	-0,0031 (-1,0709)	0,0085* (2,2568)
Kohorte 1939–1941	0,2737* (2,4406)	0,2409** (3,8064)	0,0100 (0,1158)
Kohorte 1949–1951	0,3623** (2,7425)	0,2259** (3,0529)	0,0590 (0,6162)
Verheiratet	0,5177** (3,6623)	0,4238** (5,4961)	1,6549** (17,9213)
Verheir. Männer	-0,6083** (-4,4211)	-0,3160** (-4,2415)	-1,4499** (-16,2559)
Log-Likelihood	-1887	-5182	-3139
Splits	14127	15309	14547
Ereignisse	510	1692	930

* $p \leq 0.05$, ** $p \leq 0.01$; Coeff./std. error in Klammern

Quelle: Lebensverlaufsstudie; eigene Berechnungen

Werden die diskreten Modelle basierend auf Beobachtungsintervallen von 12 Monaten hinzugenommen (Tabelle 6), so werden erste, gravierende Unterschiede deutlich. Der Parameter für die bisherige Anzahl der Beschäftigungen in dem Modell für Abwärtsmobilität ist nicht mehr signifikant. Die entsprechenden Parameter in den beiden anderen Modellen sind dagegen hoch signifikant, verändern aber ihre Vorzeichen. Gegenüber dem Modell für horizontale Mobilität in Tabelle 5 verliert der Parameter für Prestige wieder an Signifikanz. Der Parameter für das Verheiratetsein wird in den Modellen für Aufwärts- und horizontale Mobilität nun signifikant (in dem stetigen und dem diskreten Modell auf Einmonatsbasis war er es noch nicht). Insgesamt vergrößern einige Parameter ihren Abstand gegenüber Null.

In der Tabelle 7 sind die Ergebnisse für Simulationsintervalle von 24 Monaten aufgelistet. Die Parameter für Arbeitsmarkterfahrung verkleinern sich. Die Parameter für die bisherige Anzahl der Beschäftigungen zeigen wiederum gravierende Veränderungen. Gegenüber den Modellen in Tabelle 6 ist der besagte Parameter für Abwärtsmobilität nun nicht mehr signifikant. In dem Modell für horizontale Mobilität hat er sein Vorzeichen umgedreht wie ebenso in dem Modell für Aufwärtsmobilität, wo er zudem noch an Signifikanz gewonnen hat. Die Kohorte 1939–1941 zeigt gegenüber der Referenzkohorte eine höhere Rate bezüglich ihrer Aufwärtsmobilität. Die Parameter für das Verheiratetsein bleiben signifikant, sind aber nun in den Modellen für horizontale- und Abwärtsmobilität so gering, daß verheiratete Männer gegenüber Unverheirateten eine größere Neigung haben, horizontal oder abwärts mobil zu sein.

In den Ergebnissen für Beobachtungsintervalle basierend auf 48 Monaten (Tabelle 8) zeigt der Effekt für verheiratete Männer immer noch ein von den Modellen auf Monatsbasis unterschiedliches Bild. Verheiratete Männer besitzen in allen Mobilitätsrichtungen eine höhere Mobilitätsrate als Unverheiratete. Allerdings spielt Arbeitsmarkterfahrung nur noch bei horizontaler Mobilität eine Rolle. Die Anzahl der bisherigen Beschäftigungen hat nun einen durchgehend signifikanten Einfluß und die entsprechenden Parameter ein negatives Vorzeichen. Die Kohorte 1939–1941 hat jetzt nicht nur eine höhere Neigung zur horizontalen Mobilität als die Kohorte 1929–1931, sondern auch als die Kohorte 1949–1951.

Zu allerletzt sollen die Ergebnisse für Beobachtungsintervalle basierend auf einer Basis von 96 Monaten Beachtung finden (Tabelle 9). Der unplausible relative Effekt für verheiratete Männer verstärkt sich noch. Es gibt bei horizontaler Mobilität keine Unterschiede mehr bezüglich der Bildung. Arbeitsmarkterfahrung erhält ihren altbekannten Effekt zurück. Der Interzept für Aufwärtsmobilität ist nicht mehr signifikant.

Es ist hier noch anzumerken, daß zwar berufliche Auf- und Abstiege modelliert wurden, aber mit größer werdenden Beobachtungsintervallen der Verlauf der Mobilität mehr und mehr verloren geht. Werden Aufstiege be-

Tabelle 8: Competing-Risk Exponentialmodell für Simulationsintervalle von 48 Monaten

	aufwärts	horizontal	abwärts
Konstante	-0,9957** (-2,9481)	-2,2061** (-12,8176)	-2,2732** (-9,3750)
Arbeitsmarkterf.	0,0010 (1,2506)	0,0010* (2,0431)	0,0005 (0,8551)
Anzahl bish. Jobs	-0,4106** (-5,1247)	-0,5161** (-11,1040)	-0,5751** (-9,1248)
Bildung	0,1996** (7,6495)	0,0391** (2,7504)	-0,0724** (-3,5168)
Prestige	-0,1271** (-16,9696)	-0,0058 (-1,6643)	0,0107* (2,4035)
Kohorte 1939–1941	0,2707* (2,0196)	0,3619** (4,7722)	0,1180 (1,1717)
Kohorte 1949–1951	0,4447** (2,7438)	0,2690** (2,8607)	0,0865 (0,7328)
Verheiratet	0,7849** (4,6550)	0,7024** (7,3345)	1,8061** (15,6389)
Verheir. Männer	-0,5367** (-3,5262)	-0,2240** (-2,6169)	-1,3781** (-13,8856)
Log-Likelihood	-1269	-3339	-2119
Splits	7218	8079	7562
Ereignisse	374	1235	718

* $p \leq 0.05$, ** $p \leq 0.01$; Coeff./std. error in Klammern
Quelle: Lebensverlaufsstudie; eigene Berechnungen

Tabelle 9: Competing-Risk Exponentialmodell für Simulationsintervalle von 96 Monaten

	aufwärts	horizontal	abwärts
Konstante	-0,7339 (-1,5713)	-2,4679** (-9,8163)	-2,5695** (-7,7005)
Arbeitsmarkterf.	0,0060** (6,5766)	0,0067** (11,3529)	0,0068** (9,4333)
Anzahl bish. Jobs	-1,2060** (-8,1118)	-1,4151** (-15,3349)	-1,6493** (-13,2205)
Bildung	0,1627** (4,5638)	0,0153 (0,7597)	-0,0793** (-2,8361)
Prestige	-0,1291** (-13,2067)	-0,0034 (-0,7193)	0,0134* (2,3016)
Kohorte 1939–1941	0,2494 (1,4279)	0,2456* (2,3879)	0,1109 (0,8401)
Kohorte 1949–1951	0,6948** (3,1980)	0,2998* (2,2417)	0,2831 (1,7630)
Verheiratet	0,7959** (3,3646)	1,0871** (7,5293)	1,8128** (10,5020)
Verheir. Männer	-0,2169 (-1,1655)	-0,2302* (-2,1238)	-1,2320** (-9,9587)
Log-Likelihood	-730	-1748	-1184
Splits	3916	4402	4131
Ereignisse	239	725	454

* $p \leq 0.05$, ** $p \leq 0.01$; Coeff./std. error in Klammern
Quelle: Lebensverlaufsstudie; eigene Berechnungen

trachtet, so können sie, obwohl dieses vielleicht nicht mehr beobachtet werden konnte, durchaus heterogene Verläufe aufweisen. Es sind unter anderem mehrere kleine Aufstiege, ein großer Aufstieg und mehrere Auf- und Abstiege denkbar. Diese Heterogenität geht bei zunehmender Größe der Beobachtungsintervalle verloren.

4.6 Diskussion

Bei der Betrachtung des Problems der Zeitaggregation wird meistens das Short-Spell-Problem unberücksichtigt gelassen. Dieses kann insofern problematisch werden, als daß sowohl das Zeitaggregationsproblem als auch das Short-Spell-Problem Effekte aufweisen, die sich gegenseitig verstärken.

In den Beispielanalysen zeigten sich bei einer Simulation von Beobachtungsintervallen einschneidende Veränderungen in den Parameterschätzungen ereignisanalytischer Modelle. Besonders der Zeitverlauf der Rate kann von einer Zeitaggregation gravierend beeinflußt werden. Das Short-Spell-Problem zeigt dann Auswirkungen, wenn Kovariaten einen anderen Einfluß auf die längeren Episoden haben als auf die kürzeren. Einige Parameter veränderten aufgrund der Simulation ihre Vorzeichen bei zunehmender Signifikanz. Insgesamt scheinen die Folgen einer Zeitaggregation (und einer Selektion im Hinblick auf das Short-Spell-Problem) bei Parameterschätzungen schwer absehbar zu sein. Es stellt sich bei Analysen also immer die Frage, ob die Ereignisse – auch die paralleler und interdependenter Prozesse – zeitlich ausreichend genug gemessen worden sind, um ihre Geschwindigkeit und Dynamik angemessen berücksichtigen zu können.

Diskrete ereignisanalytische Modellierungen mögen von ihrer grundlegenden Konzeption hinsichtlich des Problems der Zeitaggregation als geeignetere Analyseverfahren erscheinen, doch ist diese scheinbare Angemessenheit rein oberflächlicher Natur. Weder ist aus den vorliegenden Daten abzuleiten, ob der abhängige Prozeß zeitlich genau genug erhoben wurde, noch darf die zeitliche Skalierung paralleler und interdependenter Prozesse unberücksichtigt bleiben. Prozesse können sich erheblich in ihrer Geschwindigkeit und ihrer Dynamik unterscheiden. Bei zunehmender Zeitaggregation wird es schwieriger, eine zeitliche und damit auch eine kausale Ordnung zwischen Prozessen herstellen zu können, wenn unterschiedliche Ereignisse in einem Zeitintervall stattfinden. Parameterschätzungen sind dann nicht nur von den jeweiligen Effektstärken und Einflußrichtungen zwischen Prozessen abhängig, sondern auch von der Zahl der durch Annahmen entstandenen falschen zeitlichen Ordnungen zwischen Zustandsveränderungen in einzelnen Variablen. Die zeitliche Ordnung zwischen Variablen konnte in den Beispielschätzungen kontrolliert werden (zumindest soweit es auf Monatsbasis erhobene Daten erlauben), wodurch ein auftretender Bias in den geschätzten Beispielmotellen auf die Zeitaggregation und das Short-Spell-Problem zurückgeführt werden

kann. Normalerweise kann bei stark zeitaggregierten Daten (in Abhängigkeit von der Geschwindigkeit der Prozesse) nicht immer die zeitliche und damit auch eine mögliche kausale Ordnung kontrolliert werden.

Der von Petersen (1991) vorgestellte Schätzer dürfte hinsichtlich des Zeitaggregationsproblems eher „kosmetische“ Eigenschaften besitzen, da grundlegende Probleme damit nicht gelöst werden können. Das oftmals auftretende Short-Spell-Problem bleibt unberücksichtigt und das Wichtigste: Der Bias, der sich durch die Störung des kausalen Zeitpfads ergeben kann, wird durch diesen Schätzer unberührt gelassen.

5 Zusammenfassung

Die Absicht dieses Beitrags war es, die Problematik der Zeitaggregation und des Short-Spell Problems zu verdeutlichen und mögliche Konsequenzen, die sich dadurch bei den Parameterschätzungen ergeben können, aufzuzeigen. Bei der Verwendung diskreter ereignisanalytischer Modelle als gleichsam natürliche Verfahren des Umgangs mit dem Problem der Zeitaggregation wird das oftmals zusätzlich auftretende Short-Spell-Problem ignoriert. Da bisherige Simulationen das letztere Problem unberücksichtigt ließen, wurde eine eigene Simulationsroutine entwickelt.

Bei anschließenden Modellschätzungen zeigte sich, daß in ereignisanalytischen Modellen besonders der Einfluß zeitveränderlicher Kovariaten auf die Rate sensibel auf Selektion und Zeitaggregation reagiert. Dies kann dann gravierende Auswirkungen haben, wenn bestimmte Heterogenitäten im Modell unberücksichtigt bleiben. Insgesamt bieten diskrete ereignisanalytische Modelle keinerlei Ansatzpunkt mit dem Problem der Zeitaggregation und den damit oftmals einhergehenden Problemen, Störung des Zeitpfads und Short-Spells, umzugehen.

6 Literaturverzeichnis

- AGRESTI, A. (1990). *Categorical Data Analysis*. New York u.a.: Wiley.
- ALLISON, P.D. (1982). Discrete-Time Methods for the Analysis of Event Histories. Samuel Leinhardt (Hrsg.) (1982). *Sociological Methodology 1982*. San Francisco u.a.: Jossey-Bass. S. 61–98.
- BECKER, G.S. (1975). *Human Capital*. New York: Columbia University Press.
- BECKER, G.S. (1993). *A treatise on the family*. Cambridge, M.A. u.a.: Harvard University Press.
- BERGSTRÖM, R.; EDIN, P.A. (1992). Time Aggregation and the Distributional Shape of Unemployment Duration. *Journal Of Applied Econometrics*, 7, S. 5–30.
- BLOSSFELD, H.-P. (1985). *Bildungsexpansion und Berufschancen*. Frankfurt am Main u.a.: Campus.

- BLOSSFELD, H.-P. (1992). Is the German dual system a model for a modern vocational training system? *International Journal of Comparative Sociology*, 93, S. 168–181.
- BLOSSFELD, H.-P.; HAMERLE, A.; MAYER, K.U. (1986). *Ereignisanalyse. Statistische Theorie und Anwendung in den Wirtschafts- und Sozialwissenschaften*. Frankfurt und New York: Campus.
- BLOSSFELD, H.-P.; MANTING, D.; ROHWER, G. (1993). Patterns of Change in Family Formation in the Federal Republic of Germany and the Netherlands: some Consequences for Solidarity between Generations. In Becker, H. und Hermkens, P.L.J. (Hrsg.). *Solidarity between generations. Demographic, economic, and social change, and its consequences*. (Volume I, S.175–196). Amsterdam: Thesis Publishers.
- BLOSSFELD, H.-P.; ROHWER, G. (1995). *Techniques of Event History Modeling*. Mahaw, New Jersey: Lawrence Erlbaum Associates.
- DIPRETE, T.A. (1981). Unemployment over the life cycle: Racial differences and the effects of changing economic conditions. *American Journal of Sociology*, 87, S. 286–307.
- DIGGLE, P.J.; LIANG, K.-Y.; ZEGER, S.L. (1994). *Analysis of Longitudinal Data*. Oxford: Clarendon Press.
- DÖRINGER, P.B.; PIORE, M.J. (1971). *Internal Labor Markets and Manpower Analysis*. Lexington, Massachusetts.
- KANDEL, D.B.; YAMAGUCHI, K. (1987). Job mobility and drug use: An event history analysis. *American Journal of Sociology*, 92, S. 836–878.
- KAPLAN, E.L.; MEIER, P. (1958). Nonparametric estimation from incomplete observations. *Journal of the American Statistical Association*, 53, S. 457–481.
- KELLY, J.R.; MCGRATH, J.E. (1988). *On Time and Method*. Newbury Park: Sage.
- MAYER, K.U.; BRÜCKNER, E. (1989). *Lebensverläufe und Wohlfahrtsentwicklung. Konzeption, Design und Methodik zur Erhebung von Lebensverläufen der Geburtsjahrgänge 1929–1931, 1939–1941, 1949–1951*. Teile I–III. Berlin: Max-Planck-Institut für Bildungsforschung (Materialien aus der Bildungsforschung, Nr. 35).
- PETERSEN, T. (1991). Time-Aggregation Bias in Continuous-Time Hazard-Rate Models. *Sociological Methodology 1991*. Boston, Mass. u.a.: Blackwell. S. 263–190.
- PETERSEN, T. (1993). Recent Advances in Longitudinal Methodology. *Annual Review of Sociology*, 19, S. 425–454.
- PETERSEN, T. (1995). Analysis of Event Histories. In: Arminger, G.; Clifford, C.C.; Soebel, M.E. (Hrsg.). *Handbook of Statistical Modelling for the Social and Behavioral Science*. New York u.a.: Plenum Press. S. 453–517.
- PIORE, M.J. (1969). On-the-job Training in the Dual Labor Market. Public and Private Responsibilities in the On-the-job Training of Disadvantaged Workers. In: Weber, A.R., Cassel, F.h.; Ginsburg (Hrsg.). *Public-Private Manpower Politics*. Industrial Relations Research Association Series. Madison, Wisconsin. S. 101–132.

- SØRENSEN, A.; TUMA, N.B. (1981). Labor market structures and job mobility. In: Treiman, D. und Robinson, R.V. (Hrsg.). *Research in social stratification and mobility*, Vol.1, S. 67–94. New York: Academic Press.
- STEINHAGE, N. (1999). *TACOS*. Bremen.
- STEINHAGE, N.; BLOSSFELD, H.-P. (1999). *Zur Problematik von Querschnittsdaten. Methodisch-statistische Beschränkungen von Querschnittsdaten bei der empirischen Überprüfung von Theorien*. Sfb 186 Workingpaper Nr. 62, Universität Bremen.
- TREIMANN, D.J. (1977). *Occupational prestige in comparative Perspective*. New York u.a.
- TREIMANN, D.J. (1979). Begriff und Messung des Berufsprestiges in der international vergleichenden Mobilitätsforschung. In: Pappi, F.U. (Hrsg.). *Sozialstrukturanalyse mit Umfragedaten*. Königsstein/Ts.: Athenäum. S. 124–167.
- TUMA, N.B. (1976). Rewards, resources and the rate of mobility: A nonstationary multivariate stochastic model. *American Sociological Review*, 41, S. 338–360.
- TUMA, N.B.; HANNAN, M.T. (1984). *Social dynamics. Models and methods*. New York: Academic Press.
- YAMAGUCHI, K. (1991). *Event History Analysis*. Newbury Park u.a.: Sage.